DOCUMENT RESUME

ED 411 861 IR 056 688

AUTHOR	Crane, Dennis J.
TITLE	Creating Services for the Digital Library.
PUB DATE	1996-00-00
NOTE	6p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Software; Databases; *Electronic Libraries; Futures (of Society); Information Scientists; Information Technology; Internet; Library Role; *Library Services; Online Searching; Online Systems; Reference Services; *Search Intermediaries; *Technological Advancement; *User Needs (Information); Users (Information)

ABSTRACT

The terms "virtual library," "digital library," and "electronic library" have received growing attention among professional librarians, researchers, and users of information over the past decade. The confluence of exploding sources of data, expanding technical capability, and constrained time and money will quickly move these concepts from exploration to execution by the turn of the century. Converting the potential of the digital library into reality demands a powerful array of services for users. Formulating and delivering those services will test many of the conventions under which researching and online services have operated over the past decade. This paper proposes using software and electronic content to emulate interaction between the reference librarian and the library patron as the model to define and deliver highly valued "Digital Researcher" services. These Digital Researcher services can fill a fundamental gap between the potential and the reality of digital libraries. Computers, software, and intelligent Content Bases can powerfully augment the traditional research professional. Digital Researcher services will not displace information professionals. Rather they will extend and enhance the value delivered by these professionals. Modeled after the intelligent interactions which many information professionals have daily with patrons and end-users, successful Digital Researcher services will create value, answer questions, and help expand knowledge well into the millennium. (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Creating services for the digital library

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Dennis J. Crane
Corporate & International Markets, UMI, USA

ED 411 861

Abstract: *The terms 'virtual library', 'digital library' and 'electronic library' have received growing attention among professional librarians, researchers and users of information over the past decade. The confluence of exploding sources of data, expanding technical capability and constrained time and money will quickly move these concepts from exploration to execution by the turn of the century. Converting the potential of the digital library into reality demands a powerful array of services for users. Formulating and delivering those services will test many of the conventions under which researching and online services have operated over the past decade.*
This paper proposes a focus on using software and electronic content to emulate interaction between the reference librarian and the library patron as the 'role model' to define and deliver highly valued 'Digital Researcher' services. These Digital Researcher services can fill a fundamental gap between the potential and the reality of digital libraries.

Keywords: Digital library, electronic library, natural language, services, UMI

1. Introduction

Some people go through life making things happen, some people watch things happen, and still others stand by and ask 'wha' happened?' In the world of libraries and information exchange, the confluence of technical possibility and human creativity have fundamentally altered the traditional view of the library. Many important changes are happening. These changes present great opportunity and challenge both for producers and consumers of information. These changes also alter the fundamental role of libraries in sharing information in a digital world.

As a newcomer to the field of secondary publishing and online database services, I would like to share my perspective on the transition underway in this market. I draw on more than a decade of applying information technology to change fundamentally the way organisations operate. I participated intimately in the somewhat parallel universe of business-to-business electronic commerce — the exchange of purchase orders, invoices and hundreds of other business documents from one computer application to another.

In 1984 Kenneth E. Dowlin, San Francisco's head librarian, authored a book entitled *The Electronic Library*. Slightly more than a decade later, the new San Francisco public library opened its doors and presented its patrons with a powerful example of a digitally enhanced library on a massive scale. San Francisco's library receives praise for being 'flexible enough to house computing technologies that are routinely outdated every 18 months, yet accessible enough to embody Dowlin's vision of the social role of the library' (Ref 1).

Technology has done much to capture, store, present and view a rapidly growing percentage of the world's information in digital form. Hundreds of specialised projects are advancing many aspects of the digital library. For instance, in the US the Networked Computer Science Technical Report Library (NCSTRL), pronounced 'ancestral', compiles leading edge computer research from around the world and permits Internet access. However, the project sponsors note that NCSTRL 'is not primarily about technology. It is about ... building the organisational and policy framework to support a digital library' (Ref 2).

From my perspective, the intelligent, informed, experienced reference librarian provides the greatest service value in many libraries. This implies that the service most essential to fully enabling the digital library will be 'Digital Researcher'. I would like to emphasise that the Digital Researcher will be a service, not just a technology. Some will react that technology cannot replace humans in this role. That's absolutely true. But computers, software and intelligent ContentBases can certainly and powerfully augment the traditional research professional. Technology will make many of the resources currently provided by the researcher more directly available to the user or patron.

By way of analogy, the ATM cannot entirely replace all the activities of the professionals in a branch bank. But, these devices perform many basic transactions quite reliably. More importantly ATMs provide new benefits, like 24-hour banking, that would be hard to implement in the traditional mode.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B.P. Jeapes
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

IR056688

BEST COPY AVAILABLE

2. Understanding the digital library

We can envision combining technology and processes to create a Digital Researcher service capable of interacting with users of information to deliver valuable answers to their research needs. In defining the Digital Researcher we should place important emphasis on interaction. Striving to emulate the interaction that occurs naturally between the reference librarian and the patron has many valuable implications for defining compelling services.

Developers of software and online services, researchers in computational linguistics and proponents of artificial intelligence have pursued the holy grail of posing a question to an information system, in natural language, and receiving a highly relevant answer. This sounds exciting. It also ignores an important aspect of what makes traditional reference librarians and researchers so valuable. Interaction ... the seemingly simple ability of the researcher to relate with the user or patron through a series of questions, answers, options and choices to refine a query, and then deliver the most relevant response. Successful interaction goes well beyond composing a string of text and then using some Boolean operators to make it selective. A Digital Researcher must bring much more to the party. Drawing on a large and expanding intelligent ContentBase, and constantly updated knowledge domains, a Digital Researcher would add specificity and refinement through each direct interaction with the user population.

For example, user interest in the topic 'childhood factors influencing the character of Bill White' would elicit a response from the Digital Researcher service seeking to differentiate between Bill White the business executive, Bill White the sports figure or Bill White a local politician. The original query does not suggest anything about the profession of the Bill White in question but the user clearly has one of several individuals in mind. Why not use Digital Researcher's Knowledge Domain of people and associations to hone this definition, before spinning up the searchers to chase down the 'electronic stacks', retrieving articles of questionable relevance? Isn't this after all the manner in which a professional reference librarian or researcher would interact with the user to refine a query?

3. From data to answers

We are all aware of the explosion of data in digital format. IBM reports that the cost of digital storage has declined at 40% per year for the past 20 years and projects that trend will continue (Ref 3). Statistics on the World Wide Web indicate that popular Web search engines, like Digital's Alta Vista or Lycos, have indexed more than 20 million pages of content on the Web. This pales by comparison to the hundreds of millions of pages of digital content claimed by traditional online service and CD-ROM vendors like LEXIS/NEXIS, UMI, MAID and others.

This data explosion presents both a challenge and an opportunity for digital libraries and digital researchers. Claims of being 'the biggest' database have lost their meaning. The proliferation of Internet and Web access, plus the growing number of content licensing and reseller arrangements, act to make more of the world's published data part of one huge, loosely connected database.

The explosion of data in electronic form, however, places a much greater premium on finding the right information, in the right format, at the right time, at the right price to impact key decisions and actions. This motivates greater attention to understanding 'value in use' of any particular piece of information and the effort to synthesise information on any particular topic from all the right sources to provide *answers*.

4. From publication-centric to topic-centric

Traditionally, the creation and distribution of authored content revolved around publications. Publications — newspapers, journals, periodicals and potentially News groups and listservs on the Internet — provide convenience and value added through the article selection and editorial process. Information consumers gravitated to those publications that best served their needs. Logistics and costs of print and paper distribution limited the number of sources that a typical information buyer might utilise on a regular basis.

Personally I have chosen *Business Week*, *The Wall Street Journal* and *Fortune* as publications that keep me abreast of changes in the business world. I know that *The Economist*, *Forbes*, *The Financial Times*, *Harvard Business Review* and many other publications contain equally useful content and cover many of the same topics — I just don't have the time or money to subscribe to them all.

The advent of powerful computing, software and database services has fostered a subtle but fundamental shift in information access for busy professionals and researchers. Users increasingly value the ability to survey many relevant writings on a particular topic quickly and conveniently, *regardless of source publication*. Technology and the value added services of secondary publishers and database providers tap this customer need, which is inexorably shifting the pendulum toward customised, topic-centric information delivery.

The proliferation of current awareness and news filtering services attests to the appeal of topic-centric information delivery for business users, who are inundated with potential sources of information. Publications often expand their scope and content to attract new readers and advertisers. In the process they lose some of their specific relevance for some readers. They add 'overhead' to finding the information needed to support key decisions and actions.

Database search and retrieval have long concentrated on the topic-centric needs of librarians and

researchers. Indeed, libraries are the bastion of 'topic centred' information access — one location from which patrons could retrieve most of the information they needed to address a particular issue or question. With the advent of more user-oriented information services, libraries can now provide patrons with even greater capabilities and tools. These capabilities offer access to an attractive set of relevant responses, when and where user wants to receive them.

'Disintermediation' occurs in many industries and markets when technology augments or displaces the role of the 'middle-person' in the conduct of business or the exchange of value. Experience in banking, in many areas of retailing and in telecommunications suggests that the term 're-intermediation' is more appropriate. Technology and business process change combine to define new roles and opportunities for the service intermediary. Thus large retail specialists like Toys-R-Us or Sainsbury's provide a direct link between consumer and retailer. Technology displaces traditional inventory in the supply chain. Confronted with change, creative service intermediaries including librarians, researchers and secondary publishers must create new value propositions, often for newly defined customer segments, in order to prosper.

5. The Internet and World Wide Web

Re-intermediation, enabled by technology, will continue in the publishing and information delivery industry. The decade of the 1980s and early 1990s allowed an interested few to explore electronic and digital alternatives to traditional ways of publishing, accessing and exchanging information. The popular awareness and dramatic growth of the Internet and Web over the past five years have changed all that. According to Morgan Stanley, 'the market for Internet related products and services appears to be growing more rapidly than early emerging markets for print publishing, telephony, film, radio, recorded music, television and personal computers' (Ref 4).

An incident with my 11-year-old son personalised this fundamental shift earlier this year. Seeking to help him with a school research project, I suggested that we spend a Saturday morning at the library. He informed me, in all seriousness, that he did not intend to use the library for this project. Simply put, he said 'Dad, if it's not on the Web, it's not relevant!' While far from true, this sentiment reflects how pervasive the Internet and Web have become in the publishing and information exchange arena, and it's only just begun.

6. The Digital Researcher Imperative

The digital or electronic library promises electronically to capture and house vast collections of content, covering ever more sources from which topic oriented answers can be extracted ... but only if a 'digital researcher' exists and is up to the task of locating and extracting the right answers.

Herein lies the clear imperative for providers of services to the digital library — creating, operating, maintaining and enhancing a proficient Digital Researcher that intermediates between huge collections of electronic content and a growing number of technically aware and widely connected users.

Many consider a library to be a 'place'. Technology, communications and transitory lifestyles will make 'place' much less important. The ubiquitous cash machine emulates functions of the traditional branch bank but the technology, combined with human need and creativity, goes far beyond just replacing portions of the equivalent human activity. We would think it silly to limit ATM installations to three in a row, behind a counter at the branch bank, open only from 8 am to 3 pm (or whatever your country's retail banking hours). Indeed, we find ATMs in all kinds of places, from airports to hotels to food stores, operating 24 hours per day, seven days per week.

We should fully expect that technology, need and creativity will drive the Digital Researcher far beyond replacing some *basic* functions of a reference librarian and provide greater flexibility in the process. Research for this paper, for instance, was done entirely via UMI's ProQuest Direct from office, home and hotel rooms, in cities where it would have been inconvenient to visit a conventional library. The paper was then composed on a laptop, primarily on plane trips and on a Sunday afternoon by the community swimming pool.

7. Implications for service providers

The fundamental shift toward digital libraries creates considerable opportunities for new and expanded services for the proprietors of these libraries and for the users. Focusing squarely on the Digital Researcher concept can help vendors define and deliver services that will sustain value in the new information marketplace.

7.1. Creating and delivering value

Service providers must create and deliver compelling value if they hope to secure a position in the new millennium. Service providers must foster change and address the implications of that change. In this context, the definition of value and the means of delivering value shifts quickly to reflect underlying trends in technology, organisation and economics.

For instance, technology allows more primary publishers conveniently to create, deliver and archive an electronic version of their print publications. Witness new Web sites from nearly every major news and many periodical publishers over the past two years. Web publishing encourages inclusion of graphics and images that

are missing from traditional full-text databases. Many new Web sites have a 'nearly free' pricing structure, at least for now. These realities demand that database publishers quickly move beyond traditional roles of scanning, capturing and archiving text versions of publications. They must also expand beyond traditional indexing and abstracting to add more intellectual value to the aggregated content. They must synthesise answers from a range of sources.

7.2. Powerful Digital Researcher services

In this brave new world, successful Digital Researcher providers will combine key service features in a way the users will best describe as 'powerful.' The term 'powerful' includes, but more importantly integrates, many of the characteristics frequently ascribed to leading information services today. For instance, a powerful Digital Researcher service may not include the largest database but it must access content sufficient to deliver the best answers for the selected users.

Our customers are helping UMI understand that an appealing Digital Researcher service will combine the following key components:

7.3. Digital Researcher: intelligent ContentBase

The ContentBase for the Digital Researcher must compare favourably with the resources, contacts and collections at the disposal of an experienced reference librarian or researcher. The ContentBase must be vital and must incorporate new sources in a consistent and convenient manner.

Technology can substantially augment ContentBases in several ways. Computational linguistics allows analysis and updating of terms to help keep the value-added content fresh. Frequency, relationships and other statistics help refine information requests and improve response. Living 'knowledge domains' can classify key components of the ContentBase (people, companies, industries, products). Abstracting and indexing can associate specific pieces of content with topics of interest. Adding this kind of value to the ContentBase requires that the information be sourced, loaded and keyed consistently. This criteria will be hard to achieve in the rather random world of the Web. More importantly, all of this intelligent content can and must be linked in increasingly sophisticated ways to deliver real benefit to the user.

Intelligent ContentBases will take us beyond a traditional controlled vocabulary or thesaurus. Natural language and neural networking technology continue to make progress in capturing the essence of an article and mapping the content to other articles of similar nature and content. This is but one basis for linking items in the ContentBase. Another linking opportunity stems from the potential to tie relevant Web pages conveniently to companies, products, people, advertisements and other key concepts.

An article in January's *Chronicle of Higher Education* noted 'in recent years, many works have been translated into forms computers can understand either by taking a picture of each page, or by reproducing the text itself ... pictures give researchers more of a feel for the printed page and include illustrations, but doesn't allow them to search the text for keywords ...' (Ref 5). Today and in the future, this compromise should not be required. A truly intelligent ContentBase will contain items in varying formats, including a hybrid of full text for searching and links to high resolution images relevant graphic content. Imagine, for instance, the ability to link company or product names with scanned images of related advertising or product design.

7.4. Digital Researcher: interactive interface

A highly interactive interface must complement the intelligent ContentBase in order for the Digital Researcher to deliver valuable answers to users. Again, emulating the users' interaction with an experienced reference librarian or researcher provides key design guidance for this aspect of the service.

The Digital Researcher's interaction with the user combines associations in the ContentBase with response from the user to refine the selection, synthesis and presentation of the relevant answers. Counter to some views of natural language interface, the Digital Researcher would not seek a one-shot interpretation of the users' inquiry. Rather, the Digital Researcher would use several interactions with the knowledge accumulated in the ContentBase to iterate quickly toward the right answer.

Library patrons rely on the reference librarian to know the breadth and depth of an entire set of resources and to direct the patron's inquiry. Similarly, users of the Digital Researcher will expect the system to divulge facts and suggestions that refine a query through intelligent interaction. The Digital Researcher interface must conveniently translate user interests and needs into commands and instructions to the rest of the service.

Consider an early example from UMI's ProQuest Direct Search Assistant. Most users can understand the idea of constraining a search by date range, or by specific subjects or topics. However, few care to understand the complexities of Boolean search strings or remember the syntax for a specific online service. The Search Assistant feature interacts with the user, explaining options for focusing the inquiry. Behind the scenes, these interactions are converted to Boolean logic actually to direct the search. This form of interaction continues to expand to incorporate more of the intelligence in the ContentBase as the creativity of our application designers bring more sophistication to the software.

The Digital Researcher Interface can provide important dimensions of interaction. Users should for example be able conveniently to determine the cost to retrieve specific answers that meet their needs. Straightforward interaction with an ordering module, using graphics to illustrate cost of various options, can address this need.

8. Conclusion

The provocative term 'digital library' conjures up a vision of the electronic equivalent of an institution well known to all of us. Whether in a corporate, academic or public setting, a library has four key components: the building or facility, the content collection or holdings, the services of professional librarians and researchers and last, but not the least, the end-user or patron. Digital technology has already altered the first two components fundamentally. Physical buildings and facilities decline in importance as networks and digital repositories become more powerful and less expensive. More content can be found in digital form every day without ever needing to fell a tree!.

End-users or patrons will remain in non-digital form for the foreseeable future, some *Star Trek* characters not withstanding.

This paper proposes that the missing link in the digital library remains compelling software and database services that can emulate and extend the role of the professional librarian and researcher efficiently.

Digital Researcher services will not displace information professionals, as some may fear. Rather they will extend and enhance the value delivered by these professionals. Modelled after the intelligent interactions which many in the audience have daily with your patrons and end-users, successful Digital Researcher Services will create value, answer questions and help expand knowledge well into the new millennium.

Dennis J. Crane
300 North Zeeb Road
Ann Arbor
MI 48103
USA
Tel: +1 (313) 761 4700
Fax: +1 (313) 973 9951
E-mail: dcrane@umi.com

References:

[1] Markoff, J. (1996) Data on line and on shelves: libraries of future, books from past, CyberTimes, *The New York Times*, 15 April.

[2] Anonymous (1996) Digital library system for computer scientists, *Information Today*.

[3] Grimes, W. (1996) Libraries ponder role in the digital age, *The New York Times*, A.21, 29 April.

[4] Meeker, M. and C. DePrey (1996) *The Internet Report*, Morgan Stanley, pp. 1-1 and 1-2.

[5] Wilson, D.L. (1996) Language group urges book preservation in electronic era, *The Chronicle of Higher Education*, 12 January.